REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From-to)*
30-06-2010	Final Technical Report	15-MAY-06 to 30 SEP-07

4. TITLE AND SUBTITLE
Manifold-Based Image Understanding

5a. CONTRACT NUMBER

5b. GRANT NUMBER
N00014-06-1-0829

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Baraniuk, Richard G.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Rice University
6100 Main St. MS 16
Houston TX 77005

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Reza Malek-Madani
Office of Naval Research, Code 311
875 N Randolph St.
Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)
ONR

11. SPONSORING/MONITORING AGENCY REPORT NUMBER

12. DISTRIBUTION AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This project aimed toward a unified theory and practical toolset for the analysis and processing of signal and image manifolds for signal and image understanding purposes. The unifying theme was the multiscale geometric structure of signal and image families and manifolds. Specifically, we developed theory and tools for model-based signal and image recognition and registration, sensing and compressing data on manifolds, and data-driven manifold modeling and learning. The results of our work include (1) the smashed filter, a new tool for compressive classification; (2) our theoretical proof of the sufficiency of random projections to compressively capture signals on a manifold, with application to the theory of compressive sensing; and (3) the development of new theory and algorithms for learning manifold models for signal and image families.

15. SUBJECT TERMS
compressed sensing, manifolds, dimensionality reduction, random projections, sparsity, manifold learning, image classification, object recognition, smashed filter

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE	SAR	22	Richard G. Baraniuk
U	U	U			19b. TELEPONE NUMBER (*Include area code*) 713-348-5132

Manifold-Based Image Understanding
Final Report

Richard G. Baraniuk

1 Introduction

This is the final report for ONR Grant N00014-06-1-0829 *Manifold-Based Image Understanding*. We begin by restating the motivation for our work, reviewing the project objectives, and summarizing our work. We present our results and follow each research thrust with potential future areas of work. We conclude with a list of publications supported by the grant, and a list of project personnel.

1.1 Review of motivation

The rapid growth of sensing and imaging technology combined with the need for near real-tune action based on the sensed data has rendered automatic processing, understanding, and decision making vital to our national security. The key factors complicating data acquisition and processing are many, and include:

- **Growing volumes of sensor data.** Sensor networks, UAVs and other aircraft, satellites, and other platforms operating continuously over increasingly large and complicated environments produce prodigious volumes of data that must be fused, organized, and processed, often in (near) real-time.
- **Increasingly diverse sensor data.** The apparent lack of coherency among signals and images acquired from different locations and viewpoints and with different sensor modalities and resolutions thwarts naive approaches to data fusion and processing.
- **Diverse and changing operating conditions.** Targets may be cluttered, camouflaged, or occluded and acquired or imaged under different illuminations with miscalibrated or noisy sensors. Novel targets may appear, greatly complicating online operation.

To date, these issues have been studied largely in isolation. It is our belief that real progress on signal and image understanding requires a coordinated effort based on a unified mathematical and algorithmic foundation that supports the entire processing chain, from data acquisition through processing and subsequent archiving and data mining.

Our focus in this project was on *families* of related signals and images. If one considers signals and images as points in n high-dimensional ambient space (for example, one million-pixel black-and-white digital images inhabit $\mathbb{R}^{10^6}$), then effective solutions to many signal/image/video processing problems often require exploiting the *geometric relationships* among the data in the family of interest. In many cases of practical import, the data form a lower-dimensional *manifold* structure in the high-dimensional ambient space.

1.2 Project objectives and accomplishments

This project aimed toward a unified theory and practical toolset for the analysis and processing of signal and image manifolds for signal and image understanding purposes. The unifying theme was the multiscale geometric structure of signal and image families and manifolds. Specifically, we developed theory and tools for (1) model-based signal and image recognition and registration, (2) sensing and compressing data on manifolds, and (3) data-driven manifold modeling and learning.

1. **Model-based signal recognition and registration:** We investigated how to best understand and infer signal and image information based on prior models for potential targets. Our result was the smashed filter, a new tool for compressive classification and recognition (see section 2).

2. **Sensing and compressing data on manifolds:** We sought to develop efficient sampling and measurement schemes for manifold-modeled data. Our key result was in proving the sufficiency of random projections to compressively capture signals on a manifold, and applying this result to the theory of compressive sensing (see section 3).

3. **Data-driven manifold modeling and learning:** To bridge the gap to practice, we developed new theory and algorithms for learning manifold models for signal and image families. Applications of these ideas include adaptation to novel targets, signal and image database mining, and compression (see section 4).

2 Compressive classification and recognition with the Smashed Filter

2.1 Summary of results

Compressive sensing (CS) enables the reconstruction of a sparse or compressible image or signal from a small set of linear, non-adaptive (even random) projections. However, in many applications, including object and target recognition, we are ultimately interested in making a decision about an image rather than computing a reconstruction. We have proposed a framework for *compressive classification* that operates directly on the compressive measurements without first reconstructing the image. We dub the resulting dimensionally reduced matched filter the *smashed filter*. The first part of the theory maps traditional maximum likelihood hypothesis testing into the compressive domain; we find that the number of measurements required for a given classification performance level does not depend on the sparsity or compressibility of the images but only on the noise level. The second part of the theory applies the generalized maximum likelihood method to deal with unknown transformations such as the translation, scale, or viewing angle of a target object. We exploit the fact the set of transformed images forms a low-dimensional, nonlinear manifold in the high-dimensional image space. We have found that the number of measurements required for a given classification performance level grows linearly in the dimensionality of the manifold but only logarithmically in the number of pixels/samples and image classes. Using both simulations and measurements from a new single-pixel compressive camera, we demonstrate the effectiveness of the smashed filter for target classification using very few measurements.

2.2 Background on Compressive Sensing and Classification

2.2.1 Compressive Sensing

Compressive sensing builds upon a core tenet of signal processing and information theory: that signals, images, and other data often contain some type of *structure* that enables intelligent representation and processing. As an example, many signals have a sparse representation in terms of some basis Ψ. In particular, we say that a signal $\mathbf{x} \in \mathbb{R}^N$ is K-sparse if it can be represented as $\mathbf{x} = \Psi\boldsymbol{\theta}$ where the vector $\boldsymbol{\theta} \in \mathbb{R}^N$ has only $K \ll N$ nonzero coefficients. We say that a signal is compressible if it can be closely approximated as K-sparse. The surprising result of CS is that a length-N signal that is K-sparse/compressible in some basis can be recovered exactly/approximately from a nonadaptive linear projection of the signal onto a random $O(K \log(N/K))$-dimensional basis [1, 2].

Thus we can directly acquire the signal in a compressed form.

Specifically, rather than sampling the signal, we encode $M = O(K \log(N/K))$ inner products of the signal with a set of random vectors. In matrix notation, we take the *compressive measurements*

$$\mathbf{y} = \Phi\mathbf{x}, \tag{1}$$

where $\mathbf{y}$ is an $M \times 1$ column vector, and Φ is an $M \times N$ random (but known) matrix. We will assume that Φ is an orthoprojector, i.e., that it has orthonormal rows. Since $M < N$, recovery of the signal $\mathbf{x}$ from the measurements $\mathbf{y}$ is ill-posed in general; however the additional assumption of signal sparsity or compressibility makes recovery both possible and practical.

Imaging is a particularly compelling application of CS. Our single-pixel camera [3] employs a Texas Instruments digital micromirror device (DMD), which consists of an array of N electrostatically actuated micromirrors. The camera focuses the light from the desired image onto the DMD with the mirrors set in a pseudorandom binary (0,1) pattern; the light reflecting from the mirrors set to 1 is then focused onto a single photodiode. The voltage at the photodiode is thus the inner product between the image and the random binary (0,1) pattern displayed on the DMD. Switching among M different pseudorandom patterns collects a sufficient amount of information to reconstruct or approximate the N-pixel image.

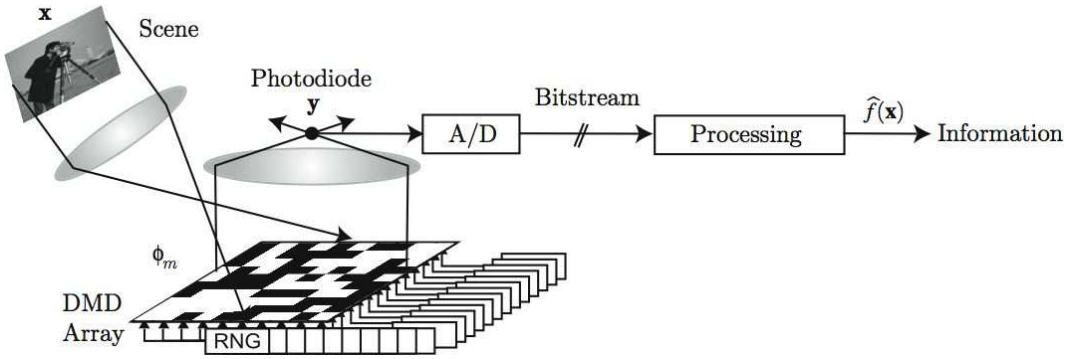

Figure 1: *Single-pixel compressive imaging camera block diagram. Incident lightfield (corresponding to the desired image* $\mathbf{x}$*) is reflected off a digital micromirror device (DMD) array whose mirror orientations are modulated in the pseudorandom pattern* ϕ_m *supplied by a random number generator. Each different mirror pattern produces a voltage at the single photodiode that corresponds to one measurement* y_m*.*

This single-pixel camera enjoys a number of advantages over traditional imaging systems. It is *universal* in the sense that the random measurements do not depend on the basis Ψ in which the image is sparse or compressible. The measurement process is also *progressive* in that better quality images can be obtained by simply taking more measurements. Compressive measurements are *democratic* in that each measurement can be given equal priority and that the reconstruction quality depends only on how many measurements are received, not on the particular subset received; this is particularly useful in remote sensing and distributed sensing where communication may be unreliable. Additionally, the use of a single detector enables imaging at new wavelengths unaccessible or prohibitively expensive using current focal plane imaging arrays.

In many applications, however, image acquisition is performed for purposes other than compression and storage. For example, in vision and surveillance settings, images are processed to extract different amounts of information from the scene observed, varying from the presence or absence of a target object to a full parametrization of the articulations of the objects in view. It is now known that compressive measurements can capture the necessary information from the image to perform such tasks [4–6]. Thus, hardware introduced for compressive imaging can also be used for these applications without modification.

2.2.2 Maximum Likelihood Classification

We begin by examining the problem of signal classification using a *maximum likelihood classifier* (MLC). Suppose a signal $\mathbf{x} \in \mathbb{R}^N$ belongs to one of P classes $\mathcal{C}_i, i = 1, \ldots, P$. We let hypothesis $\mathcal{H}_i$ signify that the signal $\mathbf{x}$ belongs to class $\mathcal{C}_i$, and for now we assume that class $\mathcal{C}_i$ contains a single known signal $\mathbf{s}_i$. We obtain noisy measurements of $\mathbf{x}$, as in $\mathbf{y} = \mathbf{x} + \boldsymbol{\omega} \in \mathbb{R}^N$, giving us a distribution $p(\mathbf{y}|\mathcal{H}_i)$ for the measured signal $\mathbf{y}$ under hypothesis $\mathcal{H}_i$. The MLC classifies according to which class has the maximum class-conditional likelihood

$$\mathcal{C}(\mathbf{y}) = \underset{i=1,\ldots,P}{\arg\max}\ p(\mathbf{y}|\mathcal{H}_i). \tag{2}$$

Under an additive white Gaussian noise (AWGN) model for $\boldsymbol{\omega}$ with variance σ, the probability distribution for the measured signal $\mathbf{y}$ under hypothesis $\mathcal{H}_i$ becomes

$$p(\mathbf{y}|\mathcal{H}_i) = \frac{1}{(2\pi\sigma)^{\frac{N}{2}}} e^{-\frac{1}{2\sigma}\|\mathbf{y}-\mathbf{s}_i\|_2^2}.$$

In this case (2) reduces to a *nearest-neighbor* classifier among the available hypotheses. Note that in the case of classification between two equally likely classes, the MLC reduces to the more common likelihood ratio test.

2.2.3 Generalized Maximum Likelihood Classification

We now consider a richer problem, where the formation of the signal $\mathbf{x}$ under each hypothesis depends on specific parameters; this results in a combined detection and estimation problem. Specifically, for each class $\mathcal{C}_i$, an element $\mathbf{x} \in \mathcal{C}_i$ can be parameterized by a unique K-dimensional parameter vector $\boldsymbol{\theta}_i \in \Theta_i$ that controls the generation of the signal, i.e., $\mathbf{x} = f_i(\boldsymbol{\theta}_i)$ for some f_i. Example parameters for image classification scenarios include the pose of the object in the scene, translation, rotation, scale, etc.

In the case of two classes, where the optimal classifier is the likelihood ratio test, we can accommodate these unknown parameters through the use of the *generalized* likelihood ratio test. We will refer to the multi-class extension of this technique as the *generalized maximum likelihood classifier* (GMLC). To derive the GMLC, we again consider the case where noisy measurements of $\mathbf{x}$ are taken, $\mathbf{y} = \mathbf{x} + \boldsymbol{\omega}$, giving us a distribution $p(\mathbf{y}|\boldsymbol{\theta}_i, \mathcal{H}_i)$ for the measured signal $\mathbf{y}$ under hypothesis $\mathcal{H}_i$ *and* parameters $\boldsymbol{\theta}_i$. The GMLC is

$$\mathcal{C}(\mathbf{y}) = \arg\max_{i=1,\ldots,P} \; p(\mathbf{y}|\widehat{\boldsymbol{\theta}}_i, \mathcal{H}_i), \tag{3}$$

where

$$\widehat{\boldsymbol{\theta}}_i = \arg\max_{\theta \in \Theta_i} \; p(\mathbf{y}|\boldsymbol{\theta}, \mathcal{H}_i) \tag{4}$$

denotes the maximum likelihood estimate (MLE) of the parameters $\boldsymbol{\theta}_i$ under hypothesis $\mathcal{H}_i$. Under the same AWGN model for $\boldsymbol{\omega}$ as above, we have

$$p(\mathbf{y}|\boldsymbol{\theta}, \mathcal{H}_i) = \frac{1}{(2\pi\sigma)^{\frac{N}{2}}} e^{-\frac{1}{2\sigma}\|\mathbf{y} - f_i(\theta)\|_2^2}, \tag{5}$$

and so the MLE (4) can be posed as

$$\widehat{\boldsymbol{\theta}}_i = \arg\min_{\theta \in \Theta_i} \; \|\mathbf{y} - f_i(\boldsymbol{\theta})\|_2^2. \tag{6}$$

2.2.4 GMLC and Smooth Manifolds

We can interpret the above situation geometrically. Consider the generation of the signal $\mathbf{x}$ under hypothesis $\mathcal{H}_i$, and recall that $\mathbf{x} = f_i(\boldsymbol{\theta}_i)$ for some K-dimensional parameter $\boldsymbol{\theta}_i \in \Theta_i$ and some function $f_i : \Theta_i \to \mathbb{R}^N$. If the mapping f_i is well-behaved, then the collection of signals $\mathcal{M}_i = \{f_i(\boldsymbol{\theta}_i) : \boldsymbol{\theta}_i \in \Theta_i\}$ forms a K-dimensional *manifold* [7] embedded in the ambient signal space $\mathbb{R}^N$. Under this interpretation, the ML estimation (6) of the parameter $\boldsymbol{\theta}_i$ under the AWGN model corresponds to finding the closest point on the manifold $\mathcal{M}_i$ to the observed signal $\mathbf{y}$. Subsequently, the classifier in (3) can be interpreted as a "nearest-manifold" search from $\mathbf{y}$ to each of the $\mathcal{M}_i$.

In settings where $\mathbf{x}$ is an image, the manifold $\mathcal{M}_i$ will have a degree of smoothness determined roughly by the spatial smoothness of the image $\mathbf{x}$ [8]. If the manifold is differentiable, then ML estimation (6) can be performed using standard optimization techniques such as Newton's method. However, the manifolds generated by images having sharp edges are not differentiable, in general. In these cases, a coarse-to-fine differential estimation procedure [8] can be used for ML estimation that involves applying Newton's method to a multiscale sequence of smoothed versions of the image.

2.2.5 The Matched Filter

We conclude the discussion of traditional classification algorithms by recalling an important special case of the GMLC—the matched filter. While it is not typically presented in this manner, the matched filter is the optimal classifier when $f_i(t; \boldsymbol{\theta}_i) = \mathbf{s}_i(t - \boldsymbol{\theta}_i)$, i.e., our observed signal is one of a known set of possible signals $\mathbf{s}_i$ shifted by an unknown amount $\boldsymbol{\theta}_i$. The matched filter simply computes $\int \mathbf{x}(t)\mathbf{s}_i(t - \boldsymbol{\theta}_i)dt$ for all possible $\boldsymbol{\theta}_i$ and selects the class with the highest peak correlation value, which is equivalent to the GMLC when the signals $\mathbf{s}_i$ have equal norm for all i and additive white Gaussian noise is assumed. This can be efficiently implemented through convolution. Due to this simplicity, matched filtering techniques are frequently applied to classification problems even when the assumptions do not necessarily hold. For instance, in some image classification settings, the classifier must be accurate but also highly efficient since the classifier may function as part of a more complex, real-time system. In these cases, despite the wide variety of more sophisticated classification algorithms, it is common to use simple matched filters for image classification, where the signals $\mathbf{s}_i$ are constructed from a set of training images and are designed so that $\langle \mathbf{x}, \mathbf{s}_i \rangle$ is large when $\mathbf{x} \in \mathcal{C}_i$ and small otherwise [9].

2.3 Compressive Classification

In this section, we formulate a classification algorithm that uses compressive measurements to exploit the low-dimensional signal manifold structure that is present in many classification applications. In some cases we know the manifold structure explicitly, while in others we learn the manifold structure from training data, which serves as a sampling of points from each of the manifolds. This structure allows us to design efficient classification systems by reducing the dimension of the data required to perform the classification.

2.3.1 Maximum Likelihood Classification

We now consider the same classification problem as in Section 2.2.2, where each class corresponds to the presence of a known signal $\mathbf{s}_i \in \mathbb{R}^N$ in noise, but instead of observing $\mathbf{x} + \boldsymbol{\omega}$ we observe $\mathbf{y} = \Phi(\mathbf{x} + \boldsymbol{\omega})$ where $\Phi \in \mathbb{R}^{M \times N}$, $M \leq N$. In this case, the MLC is essentially unchanged. In fact, when Φ is an orthoprojector and $\boldsymbol{\omega}$ is AWGN, the likelihood for each hypothesis $\mathcal{H}_i$ simply becomes

$$p(\mathbf{y}|\mathcal{H}_i) = \frac{1}{(2\pi\sigma)^{\frac{M}{2}}} e^{-\frac{1}{2\sigma}\|\mathbf{y}-\Phi\mathbf{s}_i\|_2^2}. \tag{7}$$

Thus, the MLC still reduces to nearest-neighbor classification among the available hypotheses, except in this case we compute $\|\mathbf{y} - \Phi\mathbf{s}_i\|_2$ (where $\mathbf{y} \in \mathbb{R}^M$) for each class rather than $\|\mathbf{y} - \mathbf{s}_i\|_2$ (where $\mathbf{y} \in \mathbb{R}^N$).

Two remarks are in order. First, the performance of the compressive MLC does not depend on any structure of the $\mathbf{s}_i$ such as sparsity or compressibility. Second, dimensionality reduction through a random orthoprojector will, with high probability, reduce the distance between two arbitrary points by a factor of approximately $\sqrt{M/N}$, while not affecting the variance of additive white Gaussian noise. Thus, the SNR of the projected signal is reduced compared to that in the ambient space by a factor of $\log(\sqrt{M/N})$. Hence, using the compressive MLC with $M \ll N$ may lead to an increase in the number of classification errors at high noise levels, and so M must be chosen with care [5].

2.3.2 Generalized Maximum Likelihood Classification

We are now ready to describe our GMLC-based framework for compressive classification. Assume the parameterized, multiple hypothesis setting of Section 2.2.3, and suppose that we measure $\mathbf{y} = \Phi(\mathbf{x} + \boldsymbol{\omega})$. Again assuming that Φ is an orthoprojector, the likelihood for each hypothesis $\mathcal{H}_i$ simply becomes

$$p(\mathbf{y}|\boldsymbol{\theta}, \mathcal{H}_i) = \frac{1}{(2\pi\sigma)^{\frac{M}{2}}} e^{-\frac{1}{2\sigma}\|\mathbf{y}-\Phi f_i(\theta)\|_2^2}, \tag{8}$$

which in turn reduces the MLE to

$$\widehat{\boldsymbol{\theta}}_i = \arg\min_{\theta \in \Theta_i} \|\mathbf{y} - \Phi f_i(\boldsymbol{\theta})\|_2^2. \tag{9}$$

Thus, our classifier again reduces to a "nearest-manifold" classifier; the only significant difference is that the P classes now correspond to the manifolds $\Phi\mathcal{M}_i \subset \mathbb{R}^M$, $i = 1, \ldots, P$.[1]

As above, the performance of the GMLC does not depend on any structure of the signals $\mathbf{s}_i$. Rather, its performance depends on the stability of the dimensionality reduction of the manifold: if the distances between the projected points of the manifold and the projected signal are not preserved, then the estimator performance will suffer. This issue becomes critical during the nearest-neighbor classification step.

2.3.3 Stable Embedding of Multiple Smooth Manifolds

Let us now more closely examine the random projection of one or more manifolds from a high-dimensional ambient space $\mathbb{R}^N$ to a lower-dimensional subspace $\mathbb{R}^M$. We have recently shown [7] that this process actually preserves the essential structure of a *smooth* manifold, provided that a sufficient number M of random projections are taken (see Theorem 3.1 below). Just as the CS theory demands a projection dimension M proportional to the sparsity K, the requisite M to ensure a satisfactory embedding of a manifold depends on properties of that manifold. The primary dependence is on the dimension K of the manifold, but additional factors such as the volume and curvature of the manifold play a minor role. After stating the theorem, we continue our discussion of these properties.

Theorem 1 [7] *Let $\mathcal{M}$ be a compact K-dimensional submanifold of $\mathbb{R}^N$ having condition number $1/\tau$ and volume V. Fix $0 < \epsilon < 1$ and $0 < \rho < 1$. Let Φ be a random orthoprojector from $\mathbb{R}^N$ to $\mathbb{R}^M$ with*

$$M = O\left(\frac{K \log(NV\tau^{-1}\epsilon^{-1}) \log(1/\rho)}{\epsilon^2}\right). \tag{10}$$

If $M \leq N$, then with probability at least $1 - \rho$ the following statement holds: For every pair of points $x_1, x_2 \in \mathcal{M}$,

$$(1-\epsilon)\sqrt{\frac{M}{N}} \leq \frac{\|\Phi x_1 - \Phi x_2\|_2}{\|x_1 - x_2\|_2} \leq (1+\epsilon)\sqrt{\frac{M}{N}}.$$

[1] Linear projection by Φ of a manifold $\mathcal{M} \in \mathbb{R}^N$ yields another manifold in $\Phi\mathcal{M} \in \mathbb{R}^M$.

The condition number $1/\tau$ in Theorem 3.1 is a measure of both the local curvature of the manifold and its global self-avoidance. Essentially, τ defines the maximum radius of a sphere that, when placed tangent to the manifold at any point, intersects the manifold only at that point; smaller τ leads to a larger condition number, indicating a less regular manifold [7, 10].

Theorem 3.1 states that, for sufficiently large M, all pairwise distances between points on a manifold are well-preserved under projection to a random M-dimensional subspace. (It follows also that geodesic distances are well-preserved in addition to the manifold's dimension, volume, topology, etc. [7]) By appropriate considerations of the properties $1/\tau$ and V this theorem can be extended to account for the simultaneous projection of multiple manifolds.

Corollary 1 *Let $\{\mathcal{M}_i\}_{i=1}^P$ be compact K-dimensional submanifolds of $\mathbb{R}^N$ having condition numbers $1/\tau_i$ and volumes V_i, respectively. Fix $0 < \epsilon < 1$ and $0 < \rho < 1$ and let*

$$V = \sum_i V_i \quad \text{and} \quad \tau = \min\left(\min_i \tau_i, \ \min_{i \neq j} \ \text{dist}(\mathcal{M}_i, \mathcal{M}_j)\right).$$

Let Φ be a random orthoprojector from $\mathbb{R}^N$ to $\mathbb{R}^M$ with

$$M = O\left(\frac{K \log(NV\tau^{-1}\epsilon^{-1}) \log(1/\rho)}{\epsilon^2}\right). \tag{11}$$

If $M \leq N$, then with probability at least $1-\rho$ the following statement holds: For every pair of points $x_1, x_2 \in \cup_i \mathcal{M}_i$,

$$(1-\epsilon)\sqrt{\frac{M}{N}} \leq \frac{\|\Phi x_1 - \Phi x_2\|_2}{\|x_1 - x_2\|_2} \leq (1+\epsilon)\sqrt{\frac{M}{N}}.$$

Corollary 1 ensures not only that distances between pairs of points on each manifold are well-preserved, but also that the distances between the P manifolds themselves are all well-preserved. The cost in terms of measurements is extremely modest; assuming similarly conditioned manifolds, the difference between (11) and (13) is approximately $O(K \log(P))$ additional measurements. In a classification setting with a large number of possible classes, this *sublinear* growth in the required number of measurements is particularly attractive.

Note that here the structure of the signals–namely that they inhabit low-dimensional manifolds in the high-dimensional ambient space–is critical to the performance of the GMLC. The individual signals, however, still do not need to have any sparsity or compressiblility properties; rather it is the manifolds that need to be compressible.

2.3.4 The Smashed Filter

In order to implement the GMLC described above, we first need to obtain estimates of the parameter vectors $\widehat{\boldsymbol{\theta}}_i$ from the noisy compressive measurements $\mathbf{y}$ under each of the hypotheses, which are used to classify the signal using the GMLC as described in Figure 2. We consider here some simple approaches to obtaining these estimates. Recall from Section 2.2.5 that the matched filter is an important special case of implementing the ML estimation step of the GMLC. To emphasize its similarity to the traditional matched filter while also stressing its *compressive* nature, we dub our approach the *smashed filter*.

First, suppose that we know the explicit structure of the manifold for each class, i.e., that we know the functions f_i. For example, each manifold might be parameterized by all possible shifts of a known 1-D signal or all possible positions of a known object in a 2-D image. In this setting we can explicitly search over the manifold using various approaches. In some cases we can use the same approach as the matched filter and simply continuously vary the manifold parameter(s), applying Φ to the resulting signals, and looking for the minimum distance between the output and the observed data. In the case where the original manifolds are differentiable, Corollary 1 ensures that the projected manifolds are also differentiable, and thus we can also use optimization techniques such as Newton's method to perform this optimization. In cases where the original manifolds are not differentiable, it is possible to construct measurement matrices that simultaneously smooth the data (at various resolutions) and apply a random projection [11]. This is beyond the scope of this paper but remains an interesting topic for further research.

We can also discretize an explicit search over the manifold using a grid-search procedure that, while not exhaustive, will yield a close approximation to the MLE. For example, consider the case where each class consists of a known 1-D signal $s_i(t - \boldsymbol{\theta}_i)$ with an unknown shift $\boldsymbol{\theta}_i$. In this case we can simply try a grid of possible values of $\boldsymbol{\theta}_i$ under each hypothesis assumption. This serves as a dense sampling of the manifold. Our ML estimate for $\boldsymbol{\theta}_i$

Smashed Filter Algorithm

Hypotheses:
$\mathcal{H}_i$: $\mathbf{y} = \Phi(f_i(\theta_i) + \omega)$.

Algorithm:

1. For each of the hypotheses $\mathcal{H}_1, \ldots, \mathcal{H}_P$, obtain the maximum likelihood estimate of the parameter vector

$$\widehat{\boldsymbol{\theta}}_i = \arg\min_{\theta \in \Theta_i} \ \|\mathbf{y} - \Phi f_i(\boldsymbol{\theta})\|_2^2.$$

   If $\boldsymbol{\theta}_i$ parameterizes a manifold $\mathcal{M}_i$ observed under AWGN, then $\widehat{\boldsymbol{\theta}}_i$ is the parameter value for the closest point on $\mathcal{M}_i$ to $\mathbf{y}$.

2. Perform maximum likelihood classification:

$$\mathcal{C}(\mathbf{y}) = \arg\max_{i=1,\ldots,P} \ p(\mathbf{y}|\widehat{\boldsymbol{\theta}}_i, \mathcal{H}_i),$$

   In the manifold setting with AWGN, this labels $\mathbf{y}$ with the hypothesis $\mathcal{H}_i$ for which the distance from $\mathbf{y}$ to the closest point in the manifold $\mathcal{M}_i$ is smallest.

Figure 2: Smashed filter algorithm.

for each class is simply the value that minimizes $\|y - \Phi\widehat{\mathbf{s}}_i\|_2$, where $\widehat{\mathbf{s}}_i$ denotes a sampled version of $s_i(t - \widehat{\boldsymbol{\theta}}_i)$. The classification is then based on a nearest neighbor test among the points on each manifold selected as the MLE for that class. A similar approach could be taken in any case where we know the functions f_i.

In the case where we do not explicitly know the functions f_i for each class, but rather have training data from each class, the problem simplifies even further. We can simply assume that the training data points are drawn at random from the class-appropriate manifolds. Since our classifier ultimately depends on $f_i(\widehat{\boldsymbol{\theta}}_i)$, we do not need to know the mapping f_i. We can simply estimate $f_i(\widehat{\boldsymbol{\theta}}_i)$ as the nearest-neighbor from each class in the training set. We then proceed exactly as above.

2.3.5 Advantages of Compressive Classification

In addition to the computational and storage savings afforded by compressive classification, our proposed method shares many advantages previously shown for CS reconstruction. In particular, random projections enable *universal* estimation and classification in the sense that random projections preserve the structure of any low-dimensional signal class with high probability. In our context this means that we do not need to know what the classes are or what the classification algorithm will be prior to acquiring the measurements. Additionally, compressive measurements are *progressive* in the sense that larger numbers of projections translate into higher classification rates due to increased noise tolerance and *democratic* in that each measurement can be given equal priority because classification rates depend only on how many measurements are received, not on the particular subset received.

2.4 Experimental Results

We now present results from a number of experiments that evaluate the smashed filter in a image target classification setting. We consider three classes, each for a different vehicle model: a tank, a school bus, and a truck (see Figure 3). All images are of size 128×128 pixels, hence $N = 16384$, and all measurement matrices are binary orthoprojectors obtained from a random number generator.

The first experiment is synthetic and concerns unknown shifts of a known image. In this case, $K = 2$ and we know the explicit structure of the three manifolds: each can be constructed by translating a reference image in the 2-D image plane. The shifted versions of each image, as well as their corresponding compressive measurements, were obtained synthetically using software. The first step in implementing the smashed filter is thus to find the ML estimate of the shift for each class. This can be accomplished by simply calculating the distance between the observed $\mathbf{y}$ and projections of all possible shifts of the image. For simplicity, we assume that the set of possible shifts is limited by a maximum shift of 16 pixels in any direction. After the shift estimate is obtained for each class, the GMLC selects the class whose estimate is closest to the observed image.

(a) Tank (b) School Bus (c) Truck

Figure 3: *Models used for classification experiments.*

$M = 2$			$M = 4$			$M \geq 6$		
86.1%	0.0%	13.9%	94.4%	0.0%	5.6%	100%	0.0%	0.0%
0.0%	97.2%	2.8%	0.0%	100%	0.0%	0.0%	100%	0.0%
13.9%	0.0%	86.1%	16.7%	0.0%	83.3%	0.0%	0.0%	100%

Table 1: *Confusion matrices for rotation experiments for tank, school bus, and truck with varying number of measurements M.*

We performed classification experiments for different numbers of compressive measurements, varying from $M = 2$ to 60, and for different levels of additive Gaussian noise, with standard deviations $\sigma = 0.001, 0.005, 0.01$ and 0.02. For each setting, we executed 10000 iterations of the experiment, where we selected a testing point at random with a different noise realization at each iteration. The plots of the average and minimum distances between manifolds in Figure 4(a) show that the distance is proportional roughly to the square root of the number of measurements M. In noisy measurement settings, a noise level similar to the average distance affects estimation and classification performance.

The classification rates in Figure 4(b) show a clear dependence between the number of measurements and the classification rate. Furthermore, performance for a given M degrades as the noise level increases; as expected the classifier becomes unreliable when the noise level becomes comparable to the minimum distance between the projected manifolds. In Figure 4(c), we see a similar relationship between the noise level and the error in the parameter estimate. These results verify that increasing the number of measurements improves the quality of the estimates; and the performance of the classifier is clearly dependent on the performance of the parameter estimator for the appropriate class.

The second experiment uses real data from the single-pixel camera described in Section 2.2 and concerns unknown rotations of the three targets in the z-axis in $\mathbb{R}^3$. In this case $K = 1$, and we assume that we do not know the explicit structure of the three manifolds; hence we use training data to provide an estimate of the manifold structure. We acquired a training set of compressive measurements for each vehicle for rotation angles that are multiples of 10° ($10^\circ, 20^\circ, \ldots, 360^\circ$). We first estimated the most likely rotation angle for each class by computing the nearest neighbor from each class and then performed nearest neighbor classification.

We evaluate the performance of the smashed filter classifier using leave-one-out testing. The measurements for each rotation/class combination were classified using a smashed filter trained on all other available data points. We performed classification experiments for different numbers of compressive measurements, varying from $M = 2$ to 60. Table 1 provides confusion matrices for $M = 2, 4$, and 6. The confusion matrices summarize the distributions for elements belonging to a given class (one per row: tank, school bus, or truck) being assigned a given class label (one per column: tank, school bus, or truck). The diagonal elements show the probabilities of correct classification for each of the classes. The matrices show that performance improves as M increases. Specifically, for $M \geq 6$, the classification rate remains at 100%. Figure 4(d) plots the average rotation estimate error as a function of the number of measurements, with behavior similar to that of Figure 4(b).

2.5 Conclusions and Future Work

We have demonstrated that, thanks to the pronounced structure present in many signal classes, small numbers of nonadaptive compressive measurements can suffice to capture the relevant information required for accurate classification. Our work rests on two key facts: (1) that simple parametric models impose a low-dimensional manifold structure on the signal classes within the high-dimensional image space, and (2) that the geometric structure of these manifolds is preserved under their projection to a random lower-dimensional subspace. The number of measurements required for a given classification performance level does not depend on the sparsity or compressibility of the images but only on the noise level and the structure of the manifolds, growing linearly in the

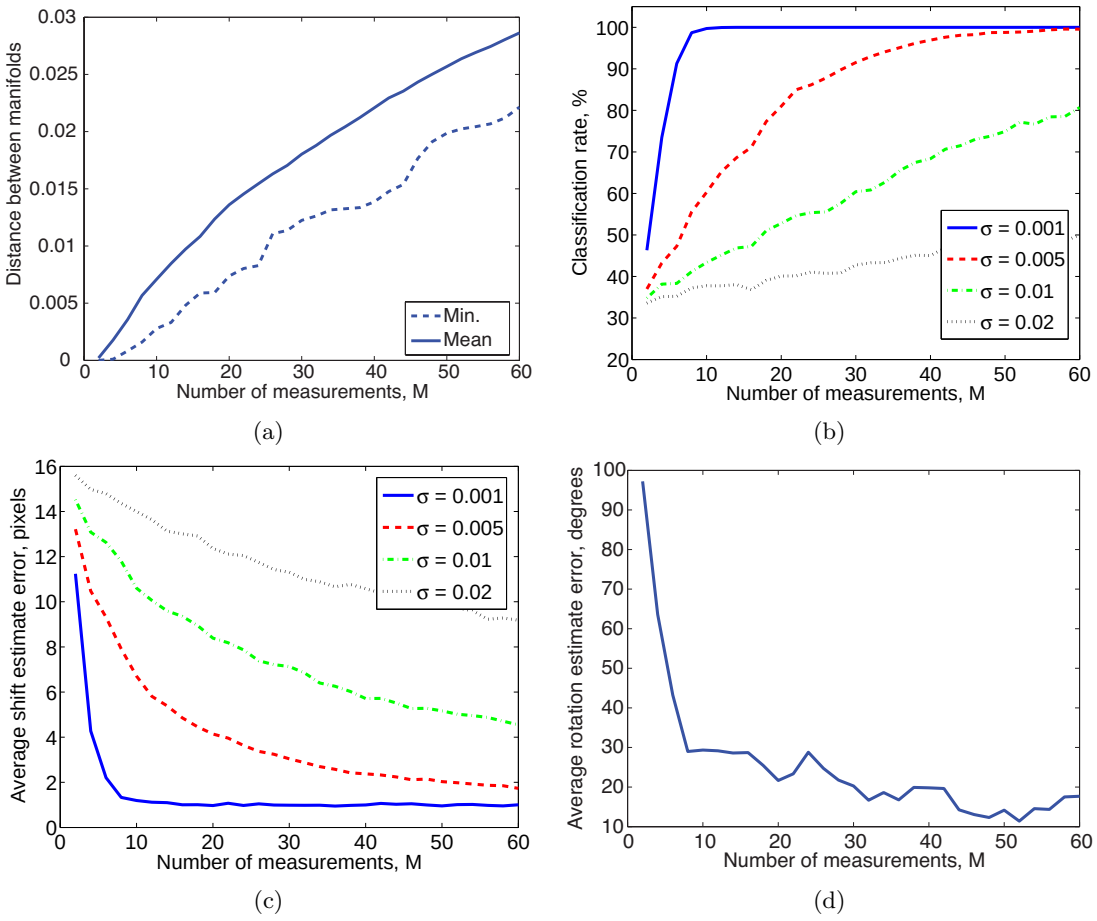

Figure 4: *Results for image classification experiments. (a) Minimum and mean distances between projected manifolds of shifted images; (b) classification rates and (c) average estimation error for varying number of measurements M and noise levels σ for the image shift experiments; (d) average estimate error for varying number of measurements M for the object rotation experiments. As M increases, the distances between the manifolds increase as well, thus increasing the noise tolerance and enabling more accurate estimation and classification. Thus, the classification and estimation performances improve as σ decreases and M increases in all cases.*

dimensionality of the manifolds but only logarithmically in the number of pixels/samples and image classes. Our GMLC-based smashed filter is readily implementable with CS hardware such as the single-pixel compressive imaging camera and shares many of the attractive features of CS in general, including simplicity, universality, robustness, democracy, and scalability, which should enable it to impact a variety of different applications.

In the future we hope to develop more sophisticated algorithms to exploit the manifold structure to more efficiently obtain the ML estimates required by the smashed filter. For example, rather than an exhaustive nearest-neighbor search, which could be computationally prohibitive for a large training set, a greedy approach might offer similar performance at significant computational savings; other approaches that exploit the smoothness of the manifolds could also be beneficial.

3 Compressed sensing on manifolds

3.1 Summary of results

We have proposed a new approach for *nonadaptive* dimensionality reduction of manifold-modeled data, demonstrating that a small number of *random linear projections* can preserve key information about a manifold-modeled signal. We center our analysis on the effect of a random linear projection operator $\Phi : \mathbb{R}^N \to \mathbb{R}^M$, $M < N$, on a smooth well-conditioned K-dimensional submanifold $\mathcal{M} \subset \mathbb{R}^N$. As our main theoretical contribution, we establish a sufficient number M of random projections to guarantee that, with high probability, all pairwise Euclidean and geodesic distances between points on $\mathcal{M}$ are well-preserved under the mapping Φ.

Our results readily admit application to the emerging theory of Compressed Sensing (CS), in which sparse signals can be recovered from small numbers of random linear measurements. As in traditional CS, the random measurements we propose can be used to recover the original data in $\mathbb{R}^N$. Moreover, like the fundamental bound in CS, our requisite M is linear in the "information level" K and logarithmic in the ambient dimension N; we also identify a logarithmic dependence on the volume and conditioning of the manifold.

3.2 Background on manifold models

Manifold signal models generalize the notion of concise signal structure beyond the framework of bases and representations. These models arise in more broad cases where (i) a K-dimensional parameter θ can be identified that carries the relevant information about a signal and (ii) the signal $x_\theta \in \mathbb{R}^N$ changes as a continuous (typically nonlinear) function of these parameters. In general, this dependence may not be neatly reflected in a sparse set of transform coefficients. Some simple explicit examples include:

- time delay of a 1-D signal (parametrized by 1 variable for translation);
- amplitude, phase, and frequency of a pure sinusoid (3 parameters);
- starting and ending time and frequency of a linear radar chirp (4 parameters);
- local signal parameters such as the configuration of a straight edge in a small image segment (2 parameters: slope and offset);
- global parameters affecting signal generation such as the position of a camera or microphone recording a scene or the relative placement of objects/speakers in a scene; and
- parameters governing the output of some articulated physical system [8, 12, 13].

In these and many other cases, the geometry of the signal class forms a nonlinear K-dimensional *submanifold* of $\mathbb{R}^N$,

$$\mathcal{F} = \{x_\theta : \theta \in \Theta\},$$

where Θ is the K-dimensional parameter space.[2] (Note the dimension K can be interpreted as an "information level" of the signal, analogous to the sparsity level K of compressive sensing.) Low-dimensional manifolds have also been proposed as approximate models for nonparametric signal classes such as images of human faces or handwritten digits [15–17].

Most algorithms for dimensionality reduction of manifold-modeled signals involve "learning" the manifold structure from a collection of data points, typically by constructing *nonlinear* mappings from $\mathbb{R}^N$ to $\mathbb{R}^M$ for some $M < N$ (ideally $M = K$) that are *adapted* to the training data and intended to preserve some characteristic property of the manifold. Example algorithms include ISOMAP [18], Hessian Eigenmaps (HLLE) [19], and Maximum Variance Unfolding (MVU) [20], which attempt to learn isometric embeddings of the manifold (preserving pairwise geodesic distances); Locally Linear Embedding (LLE) [21], which attempts to preserve local linear neighborhood structures among the embedded points; Local Tangent Space Alignment (LTSA) [22], which attempts to preserve local coordinates in each tangent space; and a method for charting a manifold [23] that attempts to preserve local neighborhood structures. These algorithms can be useful for learning the dimension and parametrizations of manifolds, for sorting data, for visualization and navigation through the data, and as preprocessing to make further analysis more tractable; common demonstrations include analysis of face images and classification of handwritten digits. A related technique, the Whitney Reduction Network [17,24], uses a training data set to adaptively construct a *linear* mapping from $\mathbb{R}^N$ to $\mathbb{R}^M$ that attempts to preserve ambient pairwise distances on the manifold; this is particularly useful for processing the output of dynamical systems having low-dimensional attractors. Additional algorithms have also been proposed for characterizing manifolds from sampled data without constructing an explicit embedding in $\mathbb{R}^M$ [10, 25, 26] and for constructing functions on the point samples in $\mathbb{R}^N$ that reflect the intrinsic structure of the manifold [27, 28]. Naturally, the burden of storing the sampled data points and implementing any of these manifold learning algorithms increases with the native dimension N of the data.

3.3 Notations and definitions

Before detailing our results, we must make clear our notation and some definitions.

[2] In general, Θ itself can be a K-dimensional manifold and need not be a subset of $\mathbb{R}^K$. We refer the reader to [14] for an excellent overview of several manifolds with relevance to signal processing, including the rotation group SO(3), which can be used for representing orientations of objects in 3-D space.

3.4 Notation

Let $d_{\mathcal{M}}(x, y)$ denote the geodesic distance between two points x and y on a K-dimensional manifold $\mathcal{M}$. Let Tan_x denote the K-dimensional tangent space to $\mathcal{M}$ at the point $x \in \mathcal{M}$. (We use the convention that Tan_x is shifted to have origin 0, rather than x.)

3.5 Condition number

To give ourselves a firm footing for analysis, we must assume a certain regularity of the manifold. For this purpose, we adopt the *condition number* defined recently by Niyogi et al. [10].

Definition 3.1 [10] *Let $\mathcal{M}$ be a compact Riemannian submanifold of $\mathbb{R}^N$. The* condition number *is defined as $1/\tau$, where τ is the largest number having the following property: The open normal bundle about $\mathcal{M}$ of radius r is embedded in $\mathbb{R}^N$ for all $r < \tau$.*

The embedding described above of the open normal bundle of radius r is a tubular neighborhood of $\mathcal{M}$ in $\mathbb{R}^N$: $\mathrm{Tub}_r = \{x + \eta \in \mathbb{R}^N : x \in \mathcal{M}, \eta \in \mathrm{Tan}_x^{\perp}, \|\eta\|_2 < r\}$, where $\mathrm{Tan}_x^{\perp}$ denotes the set of vectors normal to Tan_x. Intuitively, supposing that $r < \tau$, this tubular neighborhood does not intersect itself. More precisely, every $p \in \mathrm{Tub}_r$ corresponds to a unique $x \in \mathcal{M}$. (See also [10] for a discussion relating the condition number to the medial axis transform.)

The condition number $1/\tau$ thus controls both local properties of the manifold (such as curvature) and global properties (such as self-avoidance). The precise properties of the manifold that we require can each be derived as a consequence of the condition number. We list these explicitly below.

Lemma 3.2 [10] *(Curvature) Suppose $\mathcal{M}$ has condition number $1/\tau$. Let $p, q \in \mathcal{M}$ be two distinct points on $\mathcal{M}$, and let $\gamma(t)$ denote a unit speed parameterization of the geodesic path joining p and q. Then the curvature of γ is bounded by $1/\tau$.*

Lemma 3.3 [10] *(Twisting of tangent spaces) Suppose $\mathcal{M}$ has condition number $1/\tau$. Let $p, q \in \mathcal{M}$ be two distinct points on $\mathcal{M}$ with geodesic distance given by $d_{\mathcal{M}}(p, q)$. Let θ be the angle between the tangent spaces Tan_p and Tan_q defined by $\cos(\theta) = \min_{u \in \mathrm{Tan}_p} \max_{v \in \mathrm{Tan}_q} |\langle u, v \rangle|$. Then $\cos(\theta) > 1 - \frac{1}{\tau} d_{\mathcal{M}}(p, q)$.*

Lemma 3.4 [10] *(Self-avoidance) Suppose $\mathcal{M}$ has condition number $1/\tau$. Let $p, q \in \mathcal{M}$ be two points such that $\|p - q\|_2 = d$. Then for all $d \leq \tau/2$, the geodesic distance $d_{\mathcal{M}}(p, q)$ is bounded by $d_{\mathcal{M}}(p, q) \leq \tau - \tau\sqrt{1 - 2d/\tau}$.*

From Lemma 3.4 we have an immediate corollary.

Corollary 3.5 *Suppose $\mathcal{M}$ has condition number $1/\tau$. Let $p, q \in \mathcal{M}$ be two points such that $\|p - q\|_2 = d$. If $d \leq \tau/2$, then $d \geq d_{\mathcal{M}}(p, q) - \frac{(d_{\mathcal{M}}(p,q))^2}{2\tau}$.*

3.6 Covering regularity

We also introduce a notion of "geodesic covering regularity" for a manifold.

Definition 3.6 *Let $\mathcal{M}$ be a compact Riemannian submanifold of $\mathbb{R}^N$. Given $T > 0$, the* geodesic covering number *$G(T)$ of $\mathcal{M}$ is defined as the smallest number such that there exists a set A of points on $\mathcal{M}$, $\#A = G(T)$, so that for all $x \in \mathcal{M}$,*

$$\min_{a \in A} d_{\mathcal{M}}(x, a) \leq T.$$

Definition 3.7 *Let $\mathcal{M}$ be a compact K-dimensional Riemannian submanifold of $\mathbb{R}^N$ having volume V. We say that $\mathcal{M}$ has* geodesic covering regularity *R for resolutions $T \leq T_0$ if*

$$G(T) \leq \frac{R^K V K^{K/2}}{T^K} \tag{12}$$

for all $0 < T \leq T_0$.

The bound described in (12) reflects a natural scaling relationship as balls of radius T are used to cover a K-dimensional surface; indeed the geodesic covering regularity of the manifold is closely related to its more traditional ambient-distance covering number $C(T)$. In fact, for a manifold with condition number $1/\tau$, we can make this connection explicit. Lemma 3.4 implies that for small d, $d_{\mathcal{M}}(p,q) \leq \tau - \tau\sqrt{1-2d/\tau} \leq \tau(1-(1-2d/\tau)) = 2d$. This implies that $G(T) \leq C(T/2)$ for small T. Lemmata 5.2 and 5.3 of [10] then provide a means to bound $C(T/2)$, using the bounded curvature of the manifold to ensure that any ball centered on the manifold captures a certain amount of its volume. In particular it follows that

$$G(T) \leq C(T/2) \leq \frac{V}{\cos(\arcsin(\frac{T}{8\tau}))^K \mathrm{vol}(B^K_{T/4})} \leq \frac{V \cdot \Gamma(K/2+1)}{(1-(\frac{T}{8\tau})^2)^{K/2}\pi^{K/2}(T/4)^K} \leq \frac{3^K V K^{K/2}}{T^K}$$

for T sufficiently small (e.g., $T < \tau$). Although we point out this connection between the geodesic covering regularity and the condition number, for future reference and flexibility we prefer to specify these as distinct properties in our main result below. For notational brevity in the remainder of this paper, we neglect the minor implicit dependence of the geodesic covering regularity R on the maximum resolution T_0.

3.7 Universal dimensionality reduction for manifold models via random projections

In this paper, we propose a new approach for *nonadaptive, universal dimensionality reduction* of manifold-modeled data, demonstrating that small numbers of *random linear projections* can preserve key information about manifold-modeled signals. As we discuss, these projections can in turn be used either to recover faithful approximations to manifold-modeled signals or to discern key properties about the manifold.

We center our analysis on the effect of a random linear projection operator $\Phi : \mathbb{R}^N \to \mathbb{R}^M$ on a smooth well-conditioned K-dimensional submanifold $\mathcal{M} \subset \mathbb{R}^N$. Our main theoretical contribution (Theorem 3.1) is an RIP/JL-like result that ensures a stable embedding of the manifold in $\mathbb{R}^M$. In particular, we establish a sufficient number M of random projections to guarantee that, with high probability, all pairwise distances between points on $\mathcal{M}$ are well-preserved under the mapping Φ. Like the fundamental bound in CS, our requisite M is linear in the "information level" K and logarithmic in the ambient dimension N; additionally we identify a logarithmic dependence on the volume and conditioning of the manifold. Although the manifold itself consists of an uncountable number of points, we again exploit its low-dimensional structure to specify an effective sampling of points drawn from the manifold (plus its tangent spaces), employ the JL lemma to ensure these points are well embedded, and generalize the result to the remaining points on the manifold.

Our results suggest that, in contrast with most techniques in manifold learning, the essential information in many manifold-modeled signals can be captured via a dimensionality reducing mapping that is both *linear* and *nonadaptive*, requiring no training on sampled data and only rudimentary knowledge of $\mathcal{M}$ itself. Additionally, our results suggest that, for processing large volumes of data concentrated on manifolds, the number of requisite dimensions for a structure-preserving mapping should derive from the properties of the manifold itself, rather than the number of data points (in contrast to the JL lemma). As we discuss, these facts have promising implications in CS recovery, which can be extended beyond sparse signals to include manifold-modeled signals.

3.7.1 Theoretical result

The following result establishes a sufficient number of random projections to ensure a satisfactory embedding of a well-conditioned manifold.

Theorem 3.1 *Let $\mathcal{M}$ be a compact K-dimensional Riemannian submanifold of $\mathbb{R}^N$ having condition number $1/\tau$, volume V, and geodesic covering regularity R. Fix $0 < \epsilon < 1$ and $0 < \rho < 1$. Let Φ be a random orthoprojector from $\mathbb{R}^N$ to $\mathbb{R}^M$ with*

$$M = O\left(\frac{K \log(NVR\tau^{-1}\epsilon^{-1}) \log(1/\rho)}{\epsilon^2}\right). \tag{13}$$

If $M \leq N$, then with probability at least $1-\rho$ the following statement holds: For every pair of points $x, y \in \mathcal{M}$,

$$(1-\epsilon)\sqrt{\frac{M}{N}} \leq \frac{\|\Phi x - \Phi y\|_2}{\|x-y\|_2} \leq (1+\epsilon)\sqrt{\frac{M}{N}}. \tag{14}$$

Proof: See [7].

3.7.2 Application to Compressive Sensing

First, we consider a generalization of Compressed Sensing (CS) that takes advantage of the stable embedding of the manifold into $\mathbb{R}^M$. In the traditional CS problem, recovery of a signal $x \in \mathbb{R}^N$ from its measurement $y = \Phi x \in \mathbb{R}^M$ (where $M < N$) is possible thanks to a model that characterizes *where* in $\mathbb{R}^N$ the signal x might be expected to reside. In particular, this model comes as an assumption of sparsity in the dictionary Ψ, and the sparsity level K dictates the requisite number of measurements M to ensure stable recovery of sparse signals x. Under CS theory, recovery of sparse signals is possible because the signal set Σ_K has a stable embedding under Φ.

We have also established in this paper, however, that a signal manifold $\mathcal{M}$ can be stably embedded under a random projection from $\mathbb{R}^N$ to $\mathbb{R}^M$ for some M. This suggests that *signals obeying manifold models can also be recovered from CS measurements*, simply by replacing the traditional CS model of sparsity with a manifold model for x. Moreover, the requisite number M of measurements should depend only on the properties of the manifold $\mathcal{M}$ on which the signal x is hypothesized to live, and not on the sparsity level of x in any particular dictionary Ψ. To make this more concrete, suppose that x lives exactly on a compact submanifold $\mathcal{M} \subset \mathbb{R}^N$ with a particular dimension and condition number. Theorem 3.1 then suggests a sufficient number M of measurements so that (with high probability),[3] the operator Φ embeds $\mathcal{M}$ into $\mathbb{R}^M$. This implies also that x will be uniquely identifiable from its projection $y = \Phi x$.

Beyond simple invertibility, however, Theorem 3.1 also indicates a certain "stability" to the embedding of $\mathcal{M}$ in $\mathbb{R}^M$ (specifically, in the preservation of ambient distances). This stability can be useful for guaranteeing approximate recovery of signals that live *near* the manifold $\mathcal{M}$, rather than exactly on it. In particular, suppose $x \notin \mathcal{M}$, and let x^* be the "nearest neighbor" to x on $\mathcal{M}$, i.e.,

$$x^* = \arg \min_{x' \in \mathcal{M}} \|x - x'\|_2 , \tag{15}$$

supposing that this point is uniquely defined. Let $\widehat{x}$ be the "approximate nearest neighbor" as estimated in the projected space $\mathbb{R}^M$, i.e.,

$$\widehat{x} = \arg \min_{x' \in \mathcal{M}} \|\Phi x - \Phi x'\|_2 . \tag{16}$$

This point $\widehat{x}$ could be thought of as a "recovered" CS approximation to x. To consider this recovery successful, we would like to guarantee that $\|x - \widehat{x}\|_2$ is not much larger than $\|x - x^*\|_2$. Such a guarantee comes again from the JL lemma. Assuming that the random orthoprojector Φ is statistically independent of the signal x, then we may consider the embedding of the set $\{x\} \cup B$ under Φ. With high probability,[4] each pairwise distance in this set will have isometry constant ϵ_1. Hence, the distance from x to each anchor point will be well-preserved, and since every manifold point is no more than distance T from an anchor point, then (assuming $\|x - x^*\|_2$ is sufficiently larger than T) the distance from x to every point on $\mathcal{M}$ will be well-preserved. This guarantees a satisfactory recovery $\widehat{x}$ in the approximate nearest neighbor problem. (By examining, for example, the tangent spaces, this can all be made more precise and extended to consider the case where $\|x - x^*\|_2$ is small.)

The question remains of how a recovery program such as (16) can be efficiently solved. Unfortunately it is difficult to provide a single general-purpose algorithm, as certain nuances (such as topology) of the individual manifold can cause problem-specific challenges. (This is true even when solving (15) with the full data $x \in \mathbb{R}^N$.) However, given a suitable initial guess for the solution $\widehat{x}$ to (16), local navigation of the manifold, driven for example by an iterative Newton's method [8] to minimize $\|\Phi x - \Phi x'\|_2$, could be used to converge to the true solution. An initial guess for $\widehat{x}$ could be obtained, for example, by solving (16) using only a coarse sampling of points from $\Phi\mathcal{M}$ or by employing prior information discovered through other means. Such prior information is not unreasonable in multi-signal settings or when recovering a single signal in a multiscale fashion [31]. In general, because many key properties of the manifold are preserved, one would expect the challenge of solving (16) in the measurement space $\mathbb{R}^M$ to roughly reflect the difficulty of solving (15) in the initial ambient space $\mathbb{R}^N$.

We refer the reader to [11, 32] for additional discussion of recovery algorithms in more specific contexts such as image processing. These papers also include a series of simple but promising experiments that support manifold-based CS on a more empirical level. Additional research and experiments are ongoing, including an investigation into the relationship between the condition number and certain signal properties such as differentiability [8].

[3] Whitney's Embedding Theorem [29] actually suggests for certain K-dimensional manifolds that the number of measurements need be no larger than $2K + 1$ to ensure an embedding. However, as it offers no guarantee of stability, the practical recovery of a signal on the manifold could be complicated. Interestingly, the number $2K$ also arises in CS as the number of random measurements required to embed the set of all K-sparse signals (but again with no guarantee on the conditioning of the recovery problem); see [30].

[4] By the addition of an extra point to the embedding, there is a nominal increase in the required number of measurements. This increase becomes much more relevant in the case where a large number of signals x would need to be embedded well with respect to the manifold.

3.8 Future Work

While we have established theoretical justification for random projections capture the information of signals with a manifold structure, the question remains of how a recovery program such as (16) can be efficiently solved. Unfortunately it is difficult to provide a single general-purpose algorithm, as certain nuances (such as topology) of the individual manifold can cause problem-specific challenges. (This is true even when solving (15) with the full data $x \in \mathbb{R}^N$.) However, given a suitable initial guess for the solution $\widehat{x}$ to (16), local navigation of the manifold, driven for example by an iterative Newton's method [8] to minimize $\|\Phi x - \Phi x'\|_2$, could be used to converge to the true solution. An initial guess for $\widehat{x}$ could be obtained, for example, by solving (16) using only a coarse sampling of points from $\Phi\mathcal{M}$ or by employing prior information discovered through other means. Such prior information is not unreasonable in multi-signal settings or when recovering a single signal in a multiscale fashion [31]. In general, because many key properties of the manifold are preserved, one would expect the challenge of solving (16) in the measurement space $\mathbb{R}^M$ to roughly reflect the difficulty of solving (15) in the initial ambient space $\mathbb{R}^N$.

We refer the reader to [11, 32] for additional discussion of recovery algorithms in more specific contexts such as image processing. These papers also include a series of simple but promising experiments that support manifold-based CS on a more empirical level. Additional research and experiments are ongoing, including an investigation into the relationship between the condition number and certain signal properties such as differentiability [8].

4 Manifold learning with random projections

4.1 Summary of results

We have proposed a novel adaptive method for *linear* dimensionality reduction of manifold modeled signals. We show that with a small number of *random projections* of sample points belonging to an unknown low-dimensional Euclidean manifold, the intrinsic dimension (ID) of the sample set can be estimated up to high accuracy. Next, we rigorously prove that using only this set of random projections, we may estimate the structure of the underlying manifold. To handle practical situations, we develop a greedy-like algorithm to estimate the smallest size of the projection space required to perform manifold learning. Our method is particularly relevant in distributed sensing systems and leads to significant potential savings in data acquisition, storage and transmission costs.

4.2 Manifold learning background

An important input parameter for all manifold learning algorithms is the intrinsic dimension of a given point cloud. We would like to embed the data in as low-dimensional a space as possible, in order to avoid the familiar curse of dimensionality. However, if the embedding dimension is too small, distinct data points might get collapsed onto the same embedded point. Hence a natural question to ask would be: given a point cloud in N-dimensional Euclidean space, what is the dimension of the manifold that best captures the structure of this data set? This problem has received considerable attention in the literature and remains an active area of research [25, 33, 34].

We focus our attention on the Grassberger-Procaccia (GP) [33] algorithm for ID estimation. This is a widely used geometric[5] approach for dimensionality estimation and involves computing the *scale-dependent correlation dimension* of the data, which can be defined as follows:

Definition 4.1 *Suppose $X = (x_1, x_2, ..., x_n)$ is a finite dataset of underlying dimension K. Define*

$$C_n(r) = \frac{1}{n(n-1)} \sum_{i,j,i\neq j} \mathbf{I}_{\|x_i - x_j\| < r},$$

where $\mathbf{I}$ is the indicator function. The scale-dependent correlation dimension of X, D_{corr}, is given by

$$\widehat{D}_{\text{corr}}(r_1, r_2) = \frac{\log C_n(r_1) - \log C_n(r_2)}{\log r_1 - \log r_2}.$$

The best possible approximation to K (call this $\widehat{K}$) is obtained by fixing r_1 and r_2 as the biggest range over which the plot is linear and calculating D_{corr} in this range. There are a number of practical issues involved with this kind of approach; indeed, it has been shown that geometric ID estimation algorithms based on finite sampling yield biased estimates of intrinsic dimension [25, 35]. In our theoretical derivations, we do not attempt to take into

[5] This implies that the algorithm requires the set of pairwise distances between sample points as its input.

account the bias of the GP algorithm; instead, we prove that the effect of running the GP algorithm on a sufficient number of random projections produces a dimension estimate which well-approximates the GP estimate obtained by analyzing the original point cloud.

This estimate, $\widehat{K}$, of the intrinsic dimension of the point cloud is used by nonlinear manifold learning algorithms (e.g., Isomap [18], Locally Linear Embedding (LLE) [21] and Hessian Eigenmaps [19], among many others) to generate a $\widehat{K}$-dimensional coordinate representation of the input data points. Of these, our main analysis will be centered around the Isomap algorithm. Isomap attempts to preserve the *metric structure* of the manifold, i.e., the set of pairwise geodesic distances of any given point cloud sampled from the manifold. In essence, Isomap approximates the geodesic distances using a suitably defined graph and performs classical multidimensional scaling (MDS) to obtain a reduced K-dimensional representation of the data [18]. A key parameter in the Isomap algorithm is called the *residual variance*, which is equivalent to the stress function encountered in classical MDS. The residual variance is a measure of how well the given dataset can be embedded into a Euclidean space of dimension K. In the next section, we prescribe a specific number of measurements per data point so that performing Isomap on the randomly projected data yields a residual variance that is arbitrarily close to the variance produced by Isomap on the original dataset.

We conclude this section by revisiting the results derived in [7], which will form the basis for our results. Consider the effect of projecting a smooth K-dimensional manifold residing in $\mathbb{R}^N$ onto a random M-dimensional subspace (isomorphic to $\mathbb{R}^M$). Then, if M is sufficiently large, a stable near-isometric embedding of the manifold in the lower dimensional subspace is ensured. The key advantage is that M needs only to be *linear* in the intrinsic dimension of the manifold K. In addition, M depends only logarithmically on other properties of the manifold, such as its volume, curvature, etc. The result can be summarized in the form of the following theorem.

Theorem 4.2 [7] *Let $\mathcal{M}$ be a compact K-dimensional manifold in $\mathbb{R}^N$ having volume V and condition number $1/\tau$. Fix $0 < \epsilon < 1$ and $0 < \rho < 1$. Let Φ be a random orthoprojector[6] from $\mathbb{R}^N$ to $\mathbb{R}^M$ and*

$$M \geq O\left(\frac{K \log(NV\tau^{-1}) \log(\rho^{-1})}{\epsilon^2}\right). \tag{17}$$

Suppose $M < N$. Then, with probability exceeding $1 - \rho$, the following statement holds: For every pair of points $x, y \in \mathcal{M}$, and $i \in \{1, 2\}$,

$$(1-\epsilon)\sqrt{\frac{M}{N}} \leq \frac{d_i(\Phi x, \Phi y)}{d_i(x, y)} \leq (1+\epsilon)\sqrt{\frac{M}{N}}. \tag{18}$$

where $d_1(x, y)$ (respectively, $d_2(x, y)$) stands for the geodesic (respectively, ℓ_2) distance between points x and y.

The condition number τ controls the local, as well as global curvature of the manifold; smaller the τ, less well-conditioned the manifold with higher its "twistedness" [7]. Theorem 4.2 is proved by first specifying a finite high-resolution sampling on the manifold, the nature of which depends on its intrinsic properties; for instance, a planar manifold can be sampled coarsely. Then the Johnson-Lindenstrauss Lemma [36], a well-known result in dimensionality reduction techniques, is applied to these points to guarantee the so-called "isometry constant ϵ" (which is nothing but Equation 18).

4.3 Bounds on the performance of ID estimation and manifold learning algorithms

We note that the method of random projections essentially ensures that the metric structure of the input point cloud (i.e., the set of all pairwise distances between points belonging to the dataset) is preserved up to a distortion which depends on ϵ. This immediately suggests that geometry-based ID estimation and manifold learning algorithms could be used to study the *projected* version of the dataset.

The first of our main results establishes a sufficient number of random projections required to maintain the fidelity of the estimated correlation dimension using the GP algorithm.[7]

Theorem 4.3 *Let $\mathcal{M}$ be a compact K-dimensional manifold in $\mathbb{R}^N$ having volume V and condition number $1/\tau$. Let $X = \{x_1, x_2, ...\}$ be a sequence of samples drawn from a* uniform *density supported on $\mathcal{M}$. Let $\widehat{K}$ be the dimension*

[6] Such a matrix is formed by orthogonalizing M vectors of length N having, for example, i.i.d. Gaussian entries, assuming that they are linearly independent.

[7] The proof is detailed in a technical report provided as supplemental material.

estimate of the GP algorithm on X over the range (r_{min}, r_{max}). Let $C = 0.01$, and let $\beta = \ln(r_{max}/r_{min})$. Fix $0 < \delta < 1$ and $0 < \rho < 1$. Suppose the following condition holds:

$$\sqrt{\frac{N r_{\max}}{\tau}} < C\beta\delta. \tag{19}$$

Let Φ be a random orthoprojector from $\mathbb{R}^N$ to $\mathbb{R}^M$ with $M < N$ and

$$M \geq O\left(\frac{K \log(NV\tau^{-1}) \log(\rho^{-1})}{\beta^2 \delta^2}\right).$$

Let $\widehat{K}_\Phi$ be the estimated correlation dimension on ΦX in the projected space *over the range $(r_{min}\sqrt{M/N}, r_{max}\sqrt{M/N})$. Then, with probability exceeding $1 - \rho$, $\widehat{K}_\Phi$ is bounded by:*

$$(1 - \delta)\widehat{K} \leq \widehat{K}_\Phi \leq (1 + \delta)\widehat{K}.$$

Theorem 4.3 is a worst-case bound and serves as a sufficient condition for stable ID estimation using random projections. Thus, if we choose a sufficiently small value for δ and ρ, we are guaranteed estimation accuracy levels as close as desired to those obtained with ID estimation in the original signal space. Note that the bound on $\widehat{K}_\Phi$ is *multiplicative*. This implies that in the worst case, the number of projections required to estimate $\widehat{K}_\Phi$ very close to $\widehat{K}$ (say, within integer roundoff error) becomes higher with increasing manifold dimension K.

The second of our main theoretical results prescribes a minimum number of measurements per sample point such that the residual variance produced by Isomap in the measurement domain is within an arbitrary *additive* constant of that produced by Isomap with the full data.[8]

Theorem 4.4 *Let $\mathcal{M}$ be a compact K-dimensional manifold in $\mathbb{R}^N$ having volume V and condition number $1/\tau$. Let $X = \{x_1, x_2, ..., x_n\}$ be a finite set of samples drawn from a uniform density supported on $\mathcal{M}$. Define the* diameter Γ *of the dataset as follows:*

$$\Gamma = \max_{1 \leq i,j \leq n} d_{iso}(x_i, x_j)$$

where $d_{iso}(x, y)$ stands for the Isomap estimate of the geodesic distance between points x and y. Let Φ be a random orthoprojector from $\mathbb{R}^N$ to $\mathbb{R}^M$ with $M < N$. Fix $0 < \epsilon < 1$ and $0 < \rho < 1$. Suppose

$$M \geq O\left(\frac{K \log(NV\tau^{-1}) \log(\rho^{-1})}{\epsilon^2}\right).$$

Define R and R_Φ to be the residual variances obtained when Isomap generates a K-dimensional embedding of the original dataset X and projected dataset ΦX respectively. With probability exceeding $1 - \rho$, R_Φ is bounded by:

$$R_\Phi < R + C\Gamma^2\epsilon$$

where C is dependent only on the number of points n.

Since the choice of ϵ is arbitrary, we may choose a large enough M (which is still only logarithmic in N) such that the residual variance yielded by Isomap on the randomly projected version of the dataset is arbitrarily close to the variance produced with the native data. Again, this result is derived from a worst-case analysis. Notice that Γ acts as a measure of the scale of the dataset. In practice, we may enforce the condition that the data is normalized (i.e. every pairwise distance calculated by Isomap is divided by Γ). This ensures that the K-dimensional embedded representation is contained within a ball of unit norm centered around the origin.

Thus, we have proved that with only M measurements (where M is much smaller than the ambient dimension N) we can perform ID estimation, and subsequently learn the structure of a low-dimensional manifold, up to accuracy levels obtained by conventional methods employed in manifold learning. In Section 4.4, we utilize these sufficiency results to motivate an algorithm for performing practical manifold structure estimation using random projections.

[8] This theorem is proved in the supplemental report and relies on the proof technique used in [37].

Algorithm 1 ML-RP

$M \leftarrow 1$
$\Phi \leftarrow$ Random orthoprojector of size $M \times N$.
while residual variance $\geq \delta$ **do**
  Run the GP algorithm on ΦX.
  Use ID estimate ($\widehat{K}$) to perform Isomap on ΦX.
  Calculate residual variance.
  $M \leftarrow M + 1$
  Add one row to Φ
end while
return M
return $\widehat{K}$

4.4 How many random projections are enough?

In practice, it is hard to know, or estimate, the parameters V and τ of the underlying manifold. Also, since we have no *a priori* information of the data, it is impossible to fix $\widehat{K}$ and R, the outputs of running GP and Isomap on the original point cloud. Thus, often, we may not be able fix a definitive value for M. To circumvent this problem we develop the following empirical procedure (we call it **ML-RP**, which stands for "manifold learning using random projections".)

We initialize M to a small number, and compute M random projections of the data set $X = \{x_1, x_2, ..., x_n\}$. Using the set $\Phi X = \{\Phi x : x \in X\}$, we estimate the intrinsic dimension using the GP algorithm. This estimate, say $\widehat{K}$, is used by the Isomap algorithm to produce an embedding into $\widehat{K}$-dimensional space. The residual variance produced by this operation is recorded. For the next step of the iteration, we increment M by 1 and repeat the entire process. The algorithm terminates when the residual variance obtained is smaller than some tolerance parameter δ. A full length description is provided in Algorithm 1.

The essence of ML-RP is as follows. A sufficient number M of random projections is determined by a nonlinear procedure (i.e., sequential computation of Isomap residual variance) so that conventional manifold learning does almost as well on the projected dataset as the original. On the other hand, the random linear projections themselves provide a faithful representation of the data, in the geodesic sense. In this manner, ML-RP helps determine the number of rows that Φ requires in order to act as an operator that preserves metric structure. Therefore, ML-RP can be viewed as an adaptive method for linear reduction of data dimensionality. It is only weakly adaptive, in the sense that only the stopping criterion for ML-RP is determined by monitoring the nature of the projected data.

The results derived in Section 4.3 can be viewed as convergence proofs for ML-RP. The existence of a certain minimum number of measurements for any chosen error value δ ensures that eventually, M in the ML-RP algorithm is going to become high enough to ensure "good" Isomap performance. Also, due to the inbuilt parsimonious nature of ML-RP, it is ensured that we do not "overmeasure" the manifold, i.e., just the requisite numbers of projections of points are obtained.

4.5 Experimental results

This section details the results of simulations of ID estimation and subsequent manifold learning on real and synthetic datasets. First, we examine the performance of the GP algorithm on random projections of K-dimensional dimensional hyperspheres embedded in an ambient space of dimension $N = 150$. Figure 5(a) shows the variation of the dimension estimate produced by GP as a function of the number of projections M. The sampled dataset in each of the cases is obtained from drawing $n = 1000$ samples from a uniform distribution supported on a hypersphere of corresponding dimension. Figure 5(b) displays the minimum number of projections per sample point required to estimate the scale-dependent correlation dimension directly from the random projections, up to 10% error, when compared to GP estimation on the original data.

We observe that the ID estimate stabilizes quickly with increasing number of projections, and indeed converges to the estimate obtained by running the GP algorithm on the original data. Figure 5(b) illustrates the variation of the minimum required number of measurements M vs K, the intrinsic dimension of the underlying manifold. We plot the intrinsic dimension of the dataset against the minimum number of projections required such that $\widehat{K}_\Phi$ is within 10% of the conventional GP estimate $\widehat{K}$ (this is equivalent to choosing $\delta = 0.1$ in Theorem 4.3). We observe the predicted linearity (Theorem 4.3) in the variation of M vs K.

We turn our attention to two common datasets (Figure 6) found in the literature on dimension estimation -

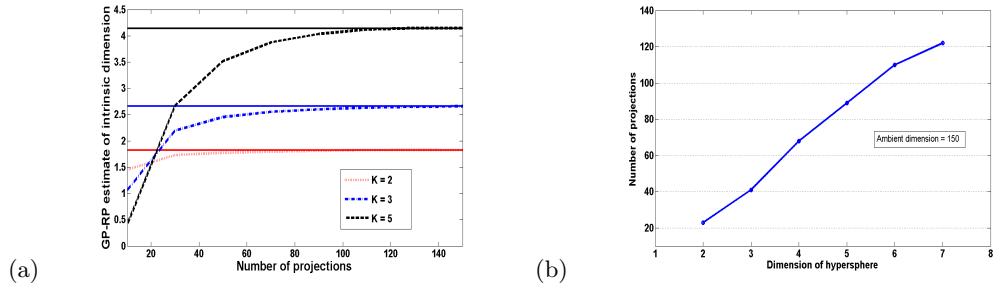

Figure 5: *Performance of ID estimation using GP as a function of random projections. Sample size $n = 1000$, ambient dimension $N = 150$. (a) Estimated intrinsic dimension for underlying hyperspherical manifolds of increasing dimension. The solid line indicates the value of the ID estimate obtained by GP performed on the original data. (b) Minimum number of projections required for GP to work with 90% accuracy as compared to GP on native data.*

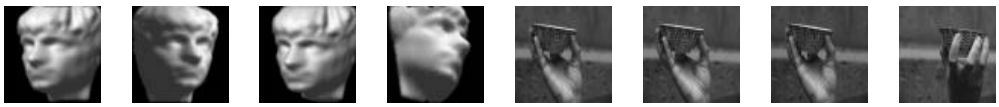

Figure 6: *Standard databases. Ambient dimension for the face database $N = 4096$; ambient dimension for the hand rotation databases $N = 3840$.*

the face database[9] [18], and the hand rotation database[10] [38]. The face database is a collection of 698 artificial snapshots of a face ($n = 64 \times 64 = 4096$) varying under 3 degrees of freedom: 2 angles for pose and 1 for lighting dimension. The signals are therefore believed to reside on a 3D manifold in an ambient space of dimension 4096. The hand rotation database is a set of 90 images ($n = 64 \times 60 = 3840$) of rotations of a hand holding an object. Although the image appearance manifold is ostensibly one-dimensional, estimators in the literature always overestimate its ID [35].

Random projections of each sample in the databases are obtained by computing the inner product of the image samples with an increasing number of rows of the random orthoprojector Φ. We notice that in the case of the face database, for $M > 60$, the Isomap variance on the randomly projected points closely approximates the variance obtained with full image data. This behavior of convergence of the variance to the best possible value is even more sharply observed in the hand rotation database, in which the two variance curves are indistinguishable for $M > 60$. These results are particularly encouraging and demonstrate the validity of the claims made in Section 4.3.

[9] *http://isomap.stanford.edu*

[10] *http://vasc.ri.cmu.edu//idb/html/motion/hand/index.html*. Note that we use a subsampled version of the database used in the literature, both in terms of resolution of the image and sampling of the manifold.

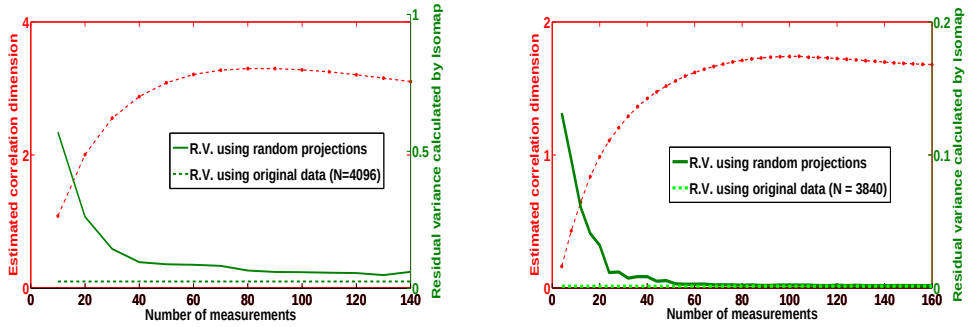

Figure 7: *Performance of ML-RP on the above databases. (left) ML-RP on the face database ($N = 4096$). Good approximates are obtained for $M > 50$. (right) ML-RP on the hand rotation database ($N = 3840$). For $M > 60$, the Isomap variance is indistinguishable from the variance obtained with full measurements.*

4.6 Conclusions and Future Work

Our main theoretical contributions are the explicit values for the lower bounds on the minimum number of random projections required to perform ID estimation, and subsequent manifold learning using Isomap, with high guaranteed accuracy levels. We propose an empirical greedy-like algorithm (ML-RP) to deal with practical situations. Experiments on simple cases, such as uniformly generated hyperspheres of varying dimension, and more complex situations such as the image databases displayed in Figure 6, have been performed and analyzed to provide sufficient evidence of the nature of the bounds described above.

The method of random projections is a powerful tool for ensuring the stable embedding of low-dimensional manifolds into a space of reasonable size. The motivation for developing results and algorithms that involve random measurements of high-dimensional data is significant, particularly due to the increasing attention that Compressed Sensing (CS) has received in the recent past. It is now possible to think of settings involving a huge number of low-power devices that inexpensively capture, store and transmit a very small number of measurements of a high-dimensional signal. ML-RP is applicable in all such situations. In situations where the bottleneck lies in the transmission of the data to the central processing node, ML-RP provides a simple solution to the manifold learning problem, and ensures that with minimum transmitted amount of information, effective manifold learning can be performed. The metric structure of the projected dataset upon termination of ML-RP closely resembles that of the original dataset with high probability; thus, ML-RP can be viewed as a novel adaptive algorithm for finding an efficient, reduced representation of data of very large dimension.

5 Publications supported by this grant

R. G. Baraniuk and M. B. Wakin, "Random projections of smooth manifolds," *Foundations of Computational Mathematics*, vol. 9, Feb. 2009.

P. T. Boufounos and R. G. Baraniuk, "Sigma delta quantization for compressive sensing," in *Wavelets XII in SPIE International Symposium on Optical Science and Technology*, (San Diego), Aug. 2007.

P. T. Boufounos, M. F. Duarte, and R. G. Baraniuk, "Sparse signal reconstruction from noisy compressive measurements using cross validation," in *Proceedings of the IEEE Workshop on Statistical Signal Processing (SSP)*, (Madison, WI), pp. 299–303, Aug. 2007.

M. A. Davenport, M. F. Duarte, M. B. Wakin, J. N. Laska, D. Takhar, K. F. Kelly, and R. G. Baraniuk, "The smashed filter for compressive classification and target recognition," in *Proceedings of Computational Imaging V at SPIE Electronic Imaging*, (San Jose, CA), Jan. 2007.

M. A. Davenport, C. Hegde, M. B. Wakin, and R. G. Baraniuk, "Manifold-based approaches for improved classification," in *Neural Information Processing Systems (NIPS) Workshop on Topology Learning*, (Whistler), Dec. 2007.

C. Hegde, M. B. Wakin, and R. G. Baraniuk, "Random projections for manifold learning," in *Neural Information Processing Systems (NIPS)*, (Vancouver), Dec. 2007.

J. N. Laska, S. Kirolos, M. F. Duarte, T. Ragheb, R. G. Baraniuk, and Y. Massoud, "Theory and implementation of an analog-to-information converter using random demodulation," in *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS)*, (New Orleans, LA), pp. 1959–1962, May 2007.

C. J. Rozell, D. H. Johnson, R. G. Baraniuk, and B. A. Olshausen, "Locally competitive algorithms for sparse approximation," in *IEEE International Conference on Image Processing – ICIP-2007*, (San Antonio, TX), Sep. 2007.

C. J. Rozell, D. H. Johnson, R. G. Baraniuk, and B. A. Olshausen, "Neurally plausible sparse coding via thresholding and local competition," *Neural Computation*, vol. 20, p. 2526–2563, Oct. 2008.

C. J. Rozell, D. H. Johnson, R. G. Baraniuk, and B. A. Olshausen, "Sparse coding via thresholding and local competition in neural circuits," *Neural computation*, vol. 20, pp. 2526–2563, Oct. 2008.

6 Professional personnel

Principal Investigator
Richard G. Baraniuk

Postdoctoral Research Associate
Petros Boufounos

Graduate Student Research Assistants
Marco Duarte, Chinmay Hegde, Jason Laska, Matthew Moravec, Christopher Rozell, Shriram Sarvotham, Mona Sheikh

References

[1] E. J. Candès and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Info. Theory*, vol. 52, pp. 5406–5425, Dec. 2006.

[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Info. Theory*, vol. 52, pp. 1289–1306, September 2006.

[3] D. Takhar, J. N. Laska, M. Wakin, M. Duarte, D. Baron, S. Sarvotham, K. K. Kelly, and R. G. Baraniuk, "A new compressive imaging camera architecture using optical-domain compression," in *Proc. IS&T/SPIE Symposium on Electronic Imaging: Computational Imaging*, vol. 6065, (San Jose, CA), pp. 43–52, Jan. 2006.

[4] M. F. Duarte, M. A. Davenport, M. B. Wakin, and R. G. Baraniuk, "Sparse signal detection from incoherent projections," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, vol. III, (Toulouse, France), pp. 305–308, May 2006.

[5] M. A. Davenport, M. B. Wakin, and R. G. Baraniuk, "Detection and estimation with compressive measurements," Tech. Rep. TREE0610, Rice University ECE Department, 2006.

[6] J. Haupt, R. Castro, R. Nowak, G. Fudge, and A. Yeh, "Compressive sampling for signal classification," in *Proc. 40th Asilomar Conf. Signals, Systems and Computers*, (Pacific Grove, CA), Oct. 2006.

[7] R. G. Baraniuk and M. B. Wakin, "Random projections of smooth manifolds," *Foundations of Computational Mathematics*, vol. 9, Feb. 2009.

[8] M. B. Wakin, D. L. Donoho, H. Choi, and R. G. Baraniuk, "The multiscale structure of non-differentiable image manifolds," in *Proc. Wavelets XI at SPIE Optics and Photonics*, (San Diego, CA), SPIE, August 2005.

[9] A. Mahalanobis, B. Kumar, and S. Sims, "Distance-classifier correlation filters for multiclass target recognition," *Applied Optics*, vol. 35, no. 17, pp. 3127–3133, 1996.

[10] P. Niyogi, S. Smale, and S. Weinberger, "Finding the homology of submanifolds with confidence from random samples," Tech. Rep. TR-2004-08, University of Chicago, 2004.

[11] M. B. Wakin, *The Geometry of Low-Dimensional Signal Models*. PhD thesis, Department of Electrical and Computer Engineering, Rice University, Houston, TX, 2006.

[12] D. L. Donoho and C. Grimes, "Image manifolds which are isometric to Euclidean space," *J. Math. Imaging Comp. Vision*, vol. 23, pp. 5–24, July 2005.

[13] C. Grimes, *New methods in nonlinear dimensionality reduction*. PhD thesis, Department of Statistics, Stanford University, 2003.

[14] I. U. Rahman, I. Drori, V. C. Stodden, D. L. Donoho, and P. Schroeder, "Multiscale representations for manifold-valued data," *SIAM J. Multiscale Model. Simul.*, vol. 4, no. 4, pp. 1201–1232, 2005.

[15] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cogn. Neurosci.*, vol. 3, no. 1, pp. 71–83, 1991.

[16] G. E. Hinton, P. Dayan, and M. Revow, "Modeling the manifolds of images of handwritten digits," *IEEE Trans. Neural Networks*, vol. 8, pp. 65–74, Jan. 1997.

[17] D. S. Broomhead and M. J. Kirby, "The Whitney Reduction Network: A method for computing autoassociative graphs," *Neural Comput.*, vol. 13, pp. 2595–2616, Nov. 2001.

[18] J. B. Tenenbaum, V. Silva, and J. C. Landford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, pp. 2319–2323, 2000.

[19] D. L. Donoho and C. E. Grimes, "Hessian Eigenmaps: Locally linear embedding techniques for high-dimensional data," *Proc. Natl. Acad. Sci. USA*, vol. 100, pp. 5591–5596, May 2003.

[20] K. Q. Weinberger and L. K. Saul, "Unsupervised learning of image manifolds by semidefinite programming," *Int. J. Comp. Vision – Special Issue: Comp. Vision Pattern Recog. (CVPR 2004)*, vol. 70, no. 1, pp. 77–90, 2006.

[21] S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, pp. 2323–2326, 2000.

[22] Z. Zhang and H. Zha, "Principal manifolds and nonlinear dimension reduction via tangent space alignment," *SIAM J. Sci. Comput.*, vol. 26, no. 1, pp. 313–338, 2005.

[23] M. Brand, "Charting a manifold," in *Advances in Neural Information Processing Systems (NIPS)*, vol. 15, pp. 985–992, MIT Press, 2003.

[24] D. S. Broomhead and M. Kirby, "A new approach for dimensionality reduction: Theory and algorithms," *SIAM J. Applied Math.*, vol. 60, no. 6, pp. 2114–2142, 2000.

[25] J. A. Costa and A. O. Hero., "Geodesic entropic graphs for dimension and entropy estimation in manifold learning," *IEEE Trans. Signal Processing*, vol. 52, pp. 2210–2221, August 2004.

[26] G. Carlsson, A. Zomorodian, A. Collins, and L. Guibas, "Persistence barcodes for shapes," in *Proc. Symp. on Geometry processing (SGP '04)*, pp. 124–135, ACM Press, 2004.

[27] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput.*, vol. 15, pp. 1373–1396, June 2003.

[28] R. R. Coifman and M. Maggioni, "Diffusion wavelets," *Appl. Comput. Harmon. Anal.*, vol. 21, pp. 53–94, June 2006.

[29] M. W. Hirsch, *Differential Topology*, vol. 33 of *Graduate Texts in Mathematics*. Springer, 1976.

[30] D. Baron, M. B. Wakin, M. F. Duarte, S. Sarvotham, and R. G. Baraniuk, "Distributed compressed sensing," 2005. Preprint.

[31] D. Donoho and Y. Tsaig, "Extensions of compressed sensing," *Signal Processing*, vol. 86, pp. 533–548, Mar. 2006.

[32] M. B. Wakin and R. G. Baraniuk, "Random projections of signal manifolds," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, IEEE, May 2006.

[33] P. Grassberger and I. Procaccia, "Measuring the strangeness of strange attractors," *Physica D Nonlinear Phenomena*, vol. 9, pp. 189–208, 1983.

[34] F. Camastra, "Data dimensionality estimation methods: a survey," *Pattern Recognition*, vol. 36, pp. 2945–2954, 2003.

[35] E. Levina and P. J. Bickel, "Maximum likelihood estimation of intrinsic dimension," in *Advances in NIPS*, vol. 17, MIT Press, 2005.

[36] S. Dasgupta and A. Gupta, "An elementary proof of the JL lemma," Tech. Rep. TR-99-006, University of California, Berkeley, 1999.

[37] M. Bernstein, V. de Silva, J. Langford, and J. Tenenbaum, "Graph approximations to geodesics on embedded manifolds." Technical report, Stanford University.

[38] B. Kégl, "Intrinsic dimension estimation using packing numbers," in *Advances in NIPS*, vol. 14, MIT Press, 2002.